

( Translation )

PATENT OFFICE  
JAPANESE GOVERNMENT

JC996 U.S. PTO

09/819179



This is to certify that the annexed is a true copy of  
the following application as filed with this Office.

#2/priority

Date of Application: September 14, 2000

Application Number: Japanese Patent Application  
No. 2000-280052

Applicant(s): Hitachi Software Engineering Co., Ltd.

January 26, 2001

Commissioner,  
Patent Office

Kozo OIKAWA (seal)

Certificate No. 2001-3002017

日 本 国 特 許 庁  
PATENT OFFICE  
JAPANESE GOVERNMENT

Jc996 U.S. PTO  
09/819179  
03/27/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
いる事項と同一であることを証明する。  
This is to certify that the annexed is a true copy of the following application as filed  
with this Office.

出 願 年 月 日  
Date of Application:

2000年 9月14日

出 願 番 号  
Application Number:

特願2000-280052

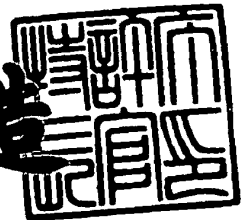
出 願 人  
Applicant(s):

日立ソフトウェアエンジニアリング株式会社

2001年 1月26日

特許庁長官  
Commissioner,  
Patent Office

及 川 耕 造



出証番号 出証特2001-3002017

【書類名】 特許願

【整理番号】 12A073

【提出日】 平成12年 9月14日

【あて先】 特許庁長官 殿

【国際特許分類】 C12N 15/00

【発明の名称】 遺伝子実験データ表示方法

【請求項の数】 8

【発明者】

【住所又は居所】 神奈川県横浜市中区尾上町6丁目81番地 日立ソフトウェアエンジニアリング株式会社内

【氏名】 中重 亮

【発明者】

【住所又は居所】 神奈川県横浜市中区尾上町6丁目81番地 日立ソフトウェアエンジニアリング株式会社内

【氏名】 野崎 康行

【発明者】

【住所又は居所】 神奈川県横浜市中区尾上町6丁目81番地 日立ソフトウェアエンジニアリング株式会社内

【氏名】 渡辺 恒彦

【発明者】

【住所又は居所】 神奈川県横浜市中区尾上町6丁目81番地 日立ソフトウェアエンジニアリング株式会社内

【氏名】 田村 卓郎

【特許出願人】

【識別番号】 000233055

【氏名又は名称】 日立ソフトウェアエンジニアリング株式会社

【代理人】

【識別番号】 100091096

【弁理士】

【氏名又は名称】 平木 祐輔

【選任した代理人】

【識別番号】 100102576

【弁理士】

【氏名又は名称】 渡辺 敏章

【先の出願に基づく優先権主張】

【出願番号】 特願2000- 86818

【出願日】 平成12年 3月27日

【手数料の表示】

【予納台帳番号】 015244

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9722155

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 遺伝子実験データ表示方法

【特許請求の範囲】

【請求項 1】 試料 A と試料 B を用いた実験で得られた試料 A と試料 B に対する複数の遺伝子の発現強度データと、試料 A と試料 C を用いた実験で得られた試料 A と試料 C に対する前記複数の遺伝子の発現強度データを、2 つの実験で共通に用いられた試料 A に対する遺伝子発現強度データを媒介として結合し、前記結合によって得られた結果を表示することを特徴とする遺伝子実験データ表示方法。

【請求項 2】 請求項 1 記載の遺伝子実験データ表示方法において、X、Y、Z 軸にそれぞれ試料 A、B、C に対する遺伝子の発現強度をとって前記複数の遺伝子の発現強度データを 3 次元表示することを特徴とする遺伝子実験データ表示方法。

【請求項 3】 請求項 2 記載の遺伝子実験データ表示方法において、各遺伝子の試料 A と試料 B に対する発現強度の比と、試料 A と試料 C に対する発現強度の比が保存されるように前記 2 つの実験における遺伝子の発現強度データを結合し、前記結合によって得られた座標位置を球面上に表示することを特徴とする遺伝子実験データ表示方法。

【請求項 4】 請求項 2 記載の遺伝子実験データ表示方法において、各遺伝子の試料 A と試料 B に対する発現強度の大小関係ならびに試料 A と試料 C に対する発現強度の大小関係が保存され、かつ各遺伝子の試料 A と試料 B に対する発現強度の比と、試料 A と試料 C に対する発現強度の比が保存されるように前記 2 つの実験における遺伝子の発現強度データを結合し、前記結合によって得られた座標位置を球体内に表示することを特徴とする遺伝子実験データ表示方法。

【請求項 5】 請求項 3 記載の遺伝子実験データ表示方法において、前記球面上に表示された各遺伝子の表示位置に基づいてクラスタ分析し、クラスタ分析の結果求められた遺伝子グループを球面上の領域として表示することを特徴とする遺伝子実験データ表示方法。

【請求項 6】 請求項 4 記載の遺伝子実験データ表示方法において、前記球

体内に表示した各遺伝子の 3 次元位置に基づいてクラスタ分析し、クラスタ分析の結果求められた遺伝子グループを球体内の領域として表示することを特徴とする遺伝子実験データ表示方法。

【請求項 7】 請求項 3 又は 4 記載の遺伝子実験データ表示方法において、前記発現強度データは時系列データであり、各遺伝子に対する各時点の発現強度データに基づく前記座標位置を時間経過の方向が分かるようにして複数表示することを特徴とする遺伝子実験データ表示方法。

【請求項 8】 請求項 5 又は 6 記載の遺伝子実験データ表示方法において、前記発現強度データは時系列データであり、前記領域の時系列変化を表示することを特徴とする遺伝子実験データ表示方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、特定の遺伝子とハイブリダイズさせることによって得られた 2 種類の実験による遺伝子発現データを、視覚的に分かりやすく、そして遺伝子の機能・役割が推測しやすい形式によって表示するための表示方法に関する。

【0002】

【従来の技術】

ゲノム配列が決定された種の増加に伴い、進化に対応すると見られる遺伝子を見つけ出し、どの生物にも共通に持っていると考えられる遺伝子の集合を探したり、それから逆に種に個別な特徴を推測するなど、種間の遺伝子の違いから何かを見出そうとする、いわゆるゲノム比較法が盛んに行われてきた。しかし近年、バイオチップ（DNA チップ）や DNA マイクロアレイなどのインフラストラクチャの発達によって、分子生物学の興味は、種間の情報から種内の情報へ、すなわち同時発生解析へと移りつつあり、これまでの種間の比較と合わせて、情報の抽出から関連付けの場が大きく広がりを持ち始めている。

【0003】

例えば、既知の遺伝子と同一の発現パターンを示す未知の遺伝子が見つければ、その遺伝子は既知の遺伝子と同様の機能があると類推できる。これら遺伝子や

蛋白質そのものの機能的な意味付けは、機能ユニットや機能グループといった形で研究されている。またそれらの間の相互作用も、既知の酵素反応データや物質代謝データとの対応付けによって、あるいはより直接的に、ある遺伝子を破壊あるいは過剰反応させ、その遺伝子の発現をなくすか、あるいは多量に発現させ、その遺伝子の直接的及び間接的影響を、全遺伝子の発現パターンを調べることによって解析している。

## 【0004】

この分野において成功した事例として、スタンフォード大学のP. Brownらのグループによるイースト菌の発現解析が挙げられる (Michel B. Eisen et al.: Cluster analysis and display of genome-wide expression patterns: Proc. Natl. Acad. Sci. (1998 Dec. 8); 95(25):pp.14863-8)。彼らは、DNAマイクロアレイを用いて、細胞から抽出した遺伝子を時系列にハイブリダイズさせ、遺伝子の発現の度合い (ハイブリダイズした蛍光シグナルの輝度) を数値化した。数値に色を対応させることで、遺伝子の個々の発現過程をわかりやすく表示させている。このとき、細胞の一連のサイクルにおいて発現パターンの過程が近い遺伝子どうし (任意の時点での発現の度合いが近いものどうし) をクラスタリングしている。

## 【0005】

また、東大医科研は薬の効力に関する実験結果を報告している (T. Tsunoda et al.: Discrimination of Drug Sensitivity of Cancer Using cDNA Microarray and Multivariate Statistical Analysis: Genome Informatics 1999 (1999, Dec.) pp.227-228, Universal Academy Press Inc.)。実験では正常細胞と癌細胞の試料が異なる色の蛍光に標識され、バイオチップ上でハイブリダイスしたのち、双方の蛍光シグナル強度が測定されている。

## 【0006】

図3は、この実験によって得られる各遺伝子の発現状態の表示方法を示した例である。この表示方法はハイブリダイズした蛍光シグナルの輝度データをプロットしたもので、片方の軸に正常細胞、他方の軸に癌細胞の輝度をとっている。データの分析においては、ある一定以上の強度のシグナルが出ている遺伝子に関し

て、正常細胞の輝度と癌細胞のシグナル強度の比に着目して、疾病に特異的な遺伝子の候補を絞り込んでいる。具体的には、図3の領域Aに入る遺伝子（正常細胞中で働いているが癌細胞中では働かなくなるもの）と領域Bに入る遺伝子（癌細胞中で働くが正常細胞中では働かなくなるもの）を特に選別している。このような表示方法をとることで、特定の疾病に特異的に働く遺伝子の候補を絞り込むことが可能となっている。

【0007】

【発明が解決しようとする課題】

図3に示した表示方法は、異なる細胞の遺伝子に関する性質の差を視覚的に大まかにとらえる際に有効であり現在一般的に利用されているが、比較したい試料は2種類までという制限がある。しかしながら、遺伝子の機能解析では多数の疾病に罹患した細胞について、それぞれの疾病に特異的な遺伝子やいくつかの疾病に特異的な遺伝子、あるいは正常な状態のみで働く遺伝子など、様々な観点から分析を行いたいという要望がある。このため、2種類に限定した表示方法は必ずしも満足できるものではない。

【0008】

例えば、正常細胞Aと疾病Pに罹った細胞Bと疾病Qに罹った細胞Cの3種類を比較したい場合は、2つずつそれぞれ実験することになる。したがって、図3のような実験結果の表示が2つ得られる。具体的には、例えば正常細胞Aと疾病Pに罹った細胞Bを対象にした実験結果の表示と、正常細胞Aと疾病Qに罹った細胞Cを対象にした実験結果の表示である。これらにより、1種類の細胞の実験値をもとにして他の2種類を比べることが可能である。しかし、図3の表示を2つ並べてみても、これら3種類の細胞における遺伝子発現がどのようなものであるか、全体のおおまかな状況を一望して判断するのは困難である。

【0009】

さらに、一般には遺伝子の発現状況を調べるために、各時点でバイオチップを使った実験を行い、遺伝子ごとに発現データの時系列を表示して複数遺伝子の変化を把握しようとする。図9はその代表的なグラフ表示例である。図9（a）は、一つの遺伝子（遺伝子1）の発現強度が時間とともにどのように変化するかを



グラフで表示したものである。また、図 9 (b) は、複数の遺伝子の時間変化を一つのグラフ上に表示したものである。図 9 (b) のグラフからは、グラフ中の丸で囲んだ部分 (9 0 0) において、遺伝子 1 と遺伝子 2 と遺伝子 3 が、ある一定の時間区間内で協調して機能していると推測できる。

#### 【0 0 1 0】

しかしながら、これらのグラフ表示方法では、前記 3 種類の細胞における発現強度の比率に着目したデータに対しては、遺伝子発現がどのように関連しているか、全体の大まかな変化の状況を判断することは困難である。

それ故本発明では、従来技術の問題点を鑑み、1 種類の遺伝子発現データをもとにして、他の複数遺伝子発現データを比較し、分類状況および変化の状況を把握するために有効なビジュアル表示を提供することを目的とする。

#### 【0 0 1 1】

##### 【課題を解決するための手段】

本発明では、前記目的を達成すべく、2 種類ごとの試料に対する遺伝子発現強度のデータを異なる実験結果の間で比較するため、共通に使用された試料のデータを媒介として 3 種類の試料に対する遺伝子発現強度を 3 次元表示する。具体的には、試料 A と試料 B に対する遺伝子発現強度データと試料 A と試料 C に対する遺伝子発現強度データを試料 A に対する遺伝子発現強度を媒介として一つの 3 次元データに変換し、球体内の点として表示する。さらにまた、試料 A と試料 B に対する遺伝子発現強度の比と試料 A と試料 C に対する遺伝子発現強度の比に着目して球面上へのマッピングを行うことで、各遺伝子の試料 A、B 及び C における発現状況を球面上の分布として表示する。これら、球体内の点の分布や球面上の点の分布をもとにクラスタ分析を行うことで、3 種類の試料における遺伝子発現状況をビジュアルに把握することが可能になる。

#### 【0 0 1 2】

すなわち、本発明による遺伝子実験データ表示方法は、試料 A と試料 B を用いた実験で得られた試料 A と試料 B に対する複数の遺伝子の発現強度データと、試料 A と試料 C を用いた実験で得られた試料 A と試料 C に対する前記複数の遺伝子の発現強度データを、2 つの実験で共通に用いられた試料 A に対する遺伝子発現

強度データを媒介として結合し、結合によって得られた結果を表示することを特徴とする。

【 0 0 1 3 】

試料 A と試料 B を用いて試料 A と試料 B に対する複数の遺伝子の発現強度データを取得する実験では、各遺伝子に試料 A と試料 B を等量混合した試料を作用させることで、試料 A と試料 B に対する実験条件を等しくするのが好ましい。同様に試料 A と試料 C を用いて試料 A と試料 C に対する複数の遺伝子の発現強度データを取得する実験では、各遺伝子に試料 A と試料 C を等量混合した試料を作用させることで、試料 A と試料 C に対する実験条件を等しくするのが好ましい。各遺伝子にハイブリダイズした試料 A と試料 B の量、あるいは試料 A と試料 C の量は、例えば試料 A と試料 B を異なる蛍光物質で標識しておき、試料 A と試料 C を異なる蛍光物質で標識しておくことで区別して検出することができる。

【 0 0 1 4 】

2 つの実験によって得られた、試料 A と試料 B に対する複数の遺伝子の発現強度データと、試料 A と試料 C に対する前記複数の遺伝子の発現強度データを、試料 A に対する遺伝子発現強度データを媒介として結合した結果の表示方法としては、例えば、試料 B に対する遺伝子発現強度を X 軸にとり、試料 C に対する遺伝子発現強度を Y 軸にとって 2 次元表示する方法がある。この表示方法によると、実際に実験を行っていない試料 B と試料 C の間における各遺伝子の発現状態を視覚表示することが可能になる。

【 0 0 1 5 】

また、2 つの実験によって得られた、試料 A と試料 B に対する複数の遺伝子の発現強度データと、試料 A と試料 C に対する前記複数の遺伝子の発現強度データを、試料 A に対する遺伝子発現強度データを媒介として結合した結果の表示方法としては、X, Y, Z 軸にそれぞれ試料 A, B, C に対する遺伝子の発現強度をとって前記複数の遺伝子の発現強度データを 3 次元表示する方法もある。この 3 次元表示方法によると、3 種類の試料に対する各遺伝子の発現状態を一目で把握することができ、類似遺伝子の分類を容易に行うことが可能になる。

【 0 0 1 6 】

遺伝子実験データを 3 次元表示する場合、各遺伝子の試料 A と試料 B に対する発現強度の比と、試料 A と試料 C に対する発現強度の比が保存されるように 2 つの実験における遺伝子の発現強度データを結合し、その結合によって得られた結果を球面上に表示するようにすることができる。さらに、球面上に表示された各遺伝子の表示位置に基づいてクラスタ分析し、クラスタ分析の結果求められた遺伝子グループを球面上の領域として表示するようにすることもできる。

#### 【 0 0 1 7 】

また、各遺伝子の試料 A と試料 B に対する発現強度の大小関係ならびに試料 A と試料 C に対する発現強度の大小関係が保存され、かつ各遺伝子の試料 A と試料 B に対する発現強度の比と、試料 A と試料 C に対する発現強度の比が保存されるように 2 つの実験における遺伝子の発現強度データを結合し、その結合によって得られた結果を球体内に表示するようにすることができる。このとき、球体は内部に位置するデータ点が見えるように半透明表示するのがよい。この場合においても、球体内に表示した各遺伝子の 3 次元位置に基づいてクラスタ分析し、クラスタ分析の結果求められた遺伝子グループを球体内の領域として表示するようにすることができる。

上記遺伝子実験データ表示方法において、発現強度データを時系列データとし、各遺伝子に対する各時点の発現強度データに基づく座標位置を時間経過の方向が分かるようにして複数表示することも有効な遺伝子実験データ表示方法である。

#### 【 0 0 1 8 】

座標位置を球面上に表示する場合には、球面上の表示位置を結ぶ矢印などによって時間経過の方向を示して表示するようにすればよい。また、座標位置を球体内に表示する場合には、球体内の 3 次元的な表示位置を結ぶ矢印などによって時間経過の方向を示して表示するようにすればよい。このように、球面上または球体内に表示された試料 A ・ 試料 B ・ 試料 C の表示位置を、時系列に従って計測した発現強度データに基づいて各遺伝子ごとに球面上または球体内に折れ線または曲線等によってトレースして軌跡を描くことにより、時間変化に対する遺伝子発現の状況を表示することが可能となる。

また、上述の遺伝子実験データ表示方法において、発現強度データを時系列データとし、クラスタ分析の結果求められた領域の時系列変化を表示することも有効な遺伝子実験データ表示方法である。

#### 【 0 0 1 9 】

領域の時系列変化は、例えば領域の移動、分割、併合などとして表示される。クラスタ分析の結果求められた遺伝子グループを球面上の領域として表示する場合には、領域の時系列変化も球面上の領域の移動、分割、併合などとして表示され、クラスタ分析の結果求められた遺伝子グループを球体内の領域として表示する場合には、領域の時系列変化も球体内の領域の移動、分割、併合などとして表示されることになる。この表示方法によると、複数遺伝子をまとめて考察する場合に、球面上または球体内に表示した遺伝子グループに対して、これを時系列に従って軌跡を計算し、遺伝子グループとしての移動・分割・合併などを描くことにより、時間変化に対する遺伝子発現の状況を表示できる。

本発明によると、複数の遺伝子発現に関する実験データを、視覚的に分かりやすく、そして遺伝子の機能・役割が推測しやすい形式によって表示することができる。

#### 【 0 0 2 0 】

##### 【発明の実施の形態】

以下、図面を参照して本発明の実施の形態を説明する。

図 1 は本発明のシステム構成図である。本発明のシステムは、一連の細胞のプロセスにおいて遺伝子の発現の度合いを数値化したものを格納している遺伝子発現データ 1 0 0、その発現データを視覚化して表示するための表示装置 1 0 1、本システムへの値の入力や選択の操作を行うためのキーボード 1 0 2 やマウス 1 0 3 などの入力装置、遺伝子の発現過程に応じて発現データのクラスタリング等の処理を行なう処理部 1 0 4 から構成される。

#### 【 0 0 2 1 】

図 2 は、本発明の遺伝子実験データ表示方法の概略処理フローを示した図である。この処理フローに従って順次説明する。

まず、遺伝子発現データ 1 0 0 から処理部 1 0 4 へデータを読み込む（ステッ

プ 2 0 0 )。図 5 に、実験結果である個々の遺伝子発現データの具体的な例を示す。図 5 は、2 つの実験の結果例を示している。実験 1 では正常の細胞 A と疾病 P に罹患した細胞 B とを比較し、実験 2 では正常の細胞と疾病 Q に罹患した細胞 C とを比較している。各実験の結果は、遺伝子 I D をインデックスとする発現量（各細胞に標識した蛍光物質から発せられる蛍光輝度の測定値）として表にまとめている。例えば実験 1 では、遺伝子 I D 1 番の遺伝子に関して、バイオチップ上のハイブリダイゼーション反応で正常細胞 A に標識した蛍光色 R の輝度が 1, 2 3 4 という数値で計測され、疾病 P に罹った細胞 B は標識した蛍光色 G の輝度が 5 6 という数値で計測された、ということを表している。対象遺伝子の総数は実験により異なるが、現在のバイオチップの実験装置としての能力に従い、数百から数万のオーダーまでを想定している。

#### 【 0 0 2 2 】

次に、遺伝子発現データの正規化処理を行う（ステップ 2 0 1）。図 5 のようなデータとして保存されている測定値を、ある 1 つの実験に着目してグラフ上にプロットすると図 3 のようになる。これは上に述べた標準的な遺伝子発現データの表示例である。しかしながら、生の蛍光輝度の測定値には、通常、機器に起因する誤差やその時に生じた実験誤差などが含まれている。

#### 【 0 0 2 3 】

そこで本実施の形態では、実験測定値の補正用としてバイオチップ上に特定の遺伝子をスポットしたものを用意する。この遺伝子はコントロールと呼ばれ、その測定値は図 3 上では白丸でグラフ上にプロットされている。図 3 では、さらにコントロールとして利用されている遺伝子が希釈されて複数スポットされ、そのデータが複数の点としてプロットされている状況を示している。この例ではコントロールの原液、1 0 分の 1 希釈液、1 0 0 分の 1 希釈液、1 0 0 0 分の 1 希釈液の 4 点が打たれている。コントロールとしては、正常細胞か否かに関わらず一定の発現量を示すことが知られている既知の遺伝子を利用するので、図 3 のグラフでこれらの点は傾き 4 5 度の直線上に希釈の度合いに応じた間隔で並ぶべきものである。したがって、実験で得られた遺伝子発現データ全体を変換し、これらのコントロールを傾き 4 5 度の直線上に並べる処理を行う。これは各点の原点の

周りの回転と拡大・縮小変換を組合わせることで行う。これらの処理を正規化処理と呼ぶ。この正規化処理を行った結果は、やはり図 5 と同様な構造をもつデータとして取得される。正規化処理後のコントロールデータは、図 4 のグラフに示すように傾き 4 5 度の直線上に並ぶ。

## 【 0 0 2 4 】

次に、正規化処理後の遺伝子発現データをプロットした図 4 のグラフにおいて、遺伝子発現データの分析対象とする範囲を設定する（ステップ 2 0 2）。測定値には輝度が小さすぎるものや大きすぎるものが含まれているので、ユーザ指定により下限の閾値 E と上限の閾値 M を設定する。これらは正規化処理後の図 4 で測定値のばらつきによって相対的に決めるものであるが、本方法では 2 つの実験結果を同時に比較して表示するため、2 つの実験について共通の値を使用する。この分析対象範囲の設定により、図 4 において斜線で示した原点付近を除く扇形の領域に入る実験値のみ、以後の表示などの処理対象となる。

## 【 0 0 2 5 】

扇形の分析対象領域に入るプロットについては 2 つの見方が可能である。一方は、正常細胞 A と疾病細胞 B のそれぞれの蛍光発現輝度の絶対量を比較して、この領域内での分布の近い・遠いといったプロットの集まり具合を判定するもので、これは、本発明では球体内にプロットして表示する表示方法に対応する。他方は、正常細胞 A と疾病細胞 B の蛍光発現輝度の相対的な大きさに着目するもので、これは例えばプロットと原点を結ぶ直線によって、図 4 の各遺伝子データに対応するプロット（黒丸の点）を大きな弧上に集め（白い四角の点）、その弧上で集まり具合を判定することと同等である。これは、本発明では球面上にプロットして表示する表示方法に対応する。

## 【 0 0 2 6 】

次に、表示を行うため、遺伝子発現データのデータ変換を行う（ステップ 2 0 3）。この変換は、各実験の発現輝度の相対比を保存する立体射影の計算に従う。具体的には以下の計算式で、各実験の測定値からそれぞれの遺伝子ごとに 3 次元表示用の座標値を求める。

## 【 0 0 2 7 】

まず、図5における実験1の測定値（正常細胞Aの測定値、疾病細胞Bの測定値）を（a, b）、実験2の測定値（正常細胞Aの測定値、疾病細胞Cの測定値）を（A, c）と記号で表すことにする。このとき、これらの2つの測定値を比の値に着目し、共通して測定した正常細胞の値が1となるようにそれぞれ変換すると、（1, b/a）及び（1, c/A）を得る。これをまとめて3次元にした測定値（b/a, c/A, 1）を用意する。次に、この3次元測定値を動径方向に拡大又は縮小し、本発明における球体表示又は球面表示とするための動径方向の拡大又は縮小量 r と R を以下の計算式で求める。ここで、 $\sqrt{\quad}$  は正の平方根を計算する関数とする。

【0028】

$$r = \sqrt{(b/a)^2 + (c/A)^2 + 1}$$

$$R = \sqrt{b^2 + c^2 + (a + A)^2}$$

この r と R を用いて、各座標値（x, y, z）は以下のように求まる。

(1) 球体内表示用座標値：（（R/r）（b/a）, （R/r）（c/A）, R/r）

(2) 球面上表示用座標値：（（K/r）（b/a）, （K/r）（c/A）, K/r）

ただし、上記の座標値中の K は、表示用球面の半径の値であり、表示する際に拡大・縮小をして見やすく調整するための定数値である。

【0029】

上記の計算によって得られた2つの座標値は、両方とも実験1および実験2で得られた発現強度の比を保存している。これは、3次元座標値をXZ平面およびYZ平面に射影することにより確認できる。2つの3次元座標で与えられる点を方向ベクトルだけに注目して、拡大または縮小のための係数 R/r または K/r を除いて考えると、3次元方向ベクトル（b/a, c/A, 1）で表される直線上にあることがわかる。これをXZ平面に射影するには y 座標値をゼロとすればよいから、射影結果として（b/a, 0, 1）という座標値を得る。これは方向ベクトルとしてみれば（b, 0, a）と同じ方向であるから、z 座標値と x 座標値の比は a 対 b になっており、実験1による正常細胞Aの測定値と疾病細胞Bの測定値の比

が保存されていることが分かる。同様にしてYZ平面に射影するにはx座標値をゼロにすればよいから、射影結果として $(0, c/A, 1)$ という座標値を得る。これは方向ベクトルとしてみれば $(0, c, A)$ と同じ方向であるから、z座標値とy座標値の比はA対cになっており、実験2による正常細胞Aの測定値と疾病細胞Cの測定値の比が保存されていることが分かる。

## 【 0 0 3 0 】

さらに、上記の計算によって得られた球体内表示用座標値は、実験1および実験2で得られた発現強度の大小関係を保存している。これは例えば、2種類の遺伝子に関する実験1の2つの測定値 $(a_1, b_1)$ と $(a_2, b_2)$ に対して、

- (1) 発現強度の比 $b_1/a_1$ と $b_2/a_2$ が同じ値、
- (2)  $a_1$ の値が $a_2$ の値より小さい（したがって $b_1$ の値も $b_2$ の値より小さい）

(3) 実験2の2つの測定値 $(A_1, c_1)$ と $(A_2, c_2)$ がほぼ同じ値、

と仮定したとき、前記球体内表示用座標値をそれぞれXZ平面に射影した点が、XZ平面内の同じ直線上にあって、さらに原点からの距離に関して $(a_1, b_1)$ に対応する射影点の方が近くにあることを意味している。実際、2つの射影点は発現強度の比が保存されるため、XZ平面内の同じ直線状にある。また原点からの距離は $R/r$ の値で決まるが、前記仮定の(1)と(3)から、2種類の遺伝子に対して計算した $r$ の値は同じになるため、その大小関係は $R$ の値で決まることになる。 $R$ の計算式によれば、前記仮定の(2)から $(a_1, b_1)$ に対応する射影点の方が近くにあることが分かる。

## 【 0 0 3 1 】

バイオチップ作成時には各スポットの量的な誤差による問題があるので、発現強度の比は、現在もっとも信頼できる定量的に測れる量である。したがって複数の実験結果を組合わせる場合にも、これらの比について定量性を保ったまま表示することが特に重要であると考えられる。本発明における球面表示に関しては、数理モデルとして2次元実射影平面となっているため、3つの実験値に関する2つの比の間の距離を球面上の大円の弧の長さで測って比べられるという利点がある。

## 【 0 0 3 2 】



発現強度の大小関係については、二つの実験を組合わせる際にバイアスが入ってしまうので、完全に定量性を保ったまま3次元表示まで変換してゆくことは難しいが、その大きい・小さいという順序は保存することが可能である。この発現強度の大小関係は、ハイブリ反応の結果である蛍光発光の強さの観測値を反映しているものであるから、定量的には対数オーダーの量的違いがあり、その大まかな違いだけでも保存しておくことは、その後のクラスタ分析による遺伝子の分類に与える影響を考慮すると、重要な意味を持つと考えられる。

## 【0033】

次に、上で求めた3次元表示用の座標値を利用して、変換後の遺伝子発現データの表示を行う（ステップ204）。すなわち、各遺伝子ごとに座標値を計算して、X、Y軸からなる直交座標系あるいはX、Y、Z軸からなる3次元直交座標系内にプロットする。これらのデータは、X座標、Y座標、Z座標が共に正の原点を中心とする8分の1の半透明球体内もしくは球面上にプロットされる。この表示状態を例示したのが図6、図7及び図8である。

## 【0034】

図6は、試料Aと試料Bの実験結果と試料Aと試料Cの実験結果とを組み合わせ、試料Bと試料Cに対する遺伝子発現強度を比較表示した例を示す図である。

これは図3と同様の表示であるが、試料Bと試料Cを直接比較するためのバイオチップを利用した実験を行っていない、という点が大きな違いである。この平面表示は、前記球体内表示用座標値の中で、x座標値とy座標値だけを部分的に取り出した平面表示用座標値： $((R/r)(b/a), (R/r)(c/A))$ を各座標値(x, y)として、XY平面にプロットしたものである。

## 【0035】

図6の解釈のしかたは図3と同様で、試料Bと試料Cの発現量の比がある一定の値以上になるプロット（領域A：直線 $Y=mX$ より上方）、あるいはある一定の値以下になるプロット（領域B：直線 $Y=(1/m)X$ より下方）が、試料Bと試料Cを比較した際に、どちらかに特異的に働く遺伝子を代表する発現データと推定することができる。

## 【0036】

図7は、半透明の球体内にプロットを行った遺伝子発現データの3次元表示例を示す図である。網掛けを施した空間領域が、図3における領域Aや領域Bなどに対応し、正常細胞A、疾病細胞B、疾病細胞Cそれぞれに特異的に働く遺伝子が観測される領域である。特に注目すべきところは、本実施の形態では疾病細胞Bと疾病細胞Cの比較実験は行っていない状況を想定しているが、試料Aと試料B、及び試料Aと試料Cの実験結果を図7のように組合わせて表示し、X-Y平面から高さHの範囲にある楔型の空間領域にプロットされる遺伝子に注目することで、疾病Bに特異的に働く遺伝子や疾病Cに特異的に働く遺伝子、あるいは共通に働く遺伝子の候補を視覚的に捕らえ選別することが可能となる。図8は、球面上にプロットを行った遺伝子発現データの表示例である。

## 【0037】

図2に戻って、最後に、表示した遺伝子発現データに対してクラスタ分析を行い、球体内もしくは球面上にそのクラスタ範囲を表示する（ステップ205）。図7及び図8には、球体内あるいは球面上にプロットを行った状況に加え、さらにクラスタ分析を行った結果を模式的に示してある。クラスタ分析は通常の変量解析で利用される階層型アルゴリズムを想定している。例えば図8に示した球面上にプロットする表示方法の場合、2つのプロット間の距離は、もとの2つの実験の結果から保存されている発現量比に基づき、球面上の距離（2点間を繋ぐ大円の弧の長さ）で計るものとする。クラスタ併合のアルゴリズムは、ウォード法・群平均法・最短距離法・最長距離法・重心法・メディアン法など、どの方法を採用してもよい。図8では、正常細胞では働いていないが疾病細胞Bで特異的に機能する遺伝子のクラスタVや疾病細胞Cで特異的に機能する遺伝子のクラスタWが検出された状況を示している。

## 【0038】

以上のように図2に示した処理フローに従うことによって、試料Aと試料B、及び試料Aと試料Cの実験結果から図6に示したような試料Bと試料Cに対する遺伝子発現強度の比較表示、あるいは図7や図8に示したような遺伝子発現データの球体内もしくは球面上のプロット、及びクラスタ分析結果の表示が可能とな

る。また、前記 3 次元表示においてはユーザの視点方向ををマウスなどにより指定し、X Y Z 座標系を回転したり位置をずらしたりすることで、検出されたクラスタ見やすい位置に配置して確認することも可能である。

## 【 0 0 3 9 】

さらに、遺伝子の時間変化に伴う発現状況を調べるために、各時点でバイオチップなどを使った実験行い、遺伝子ごとに発現データの時系列を表示して複数遺伝子の変化を把握しようとする場合、前記の球面上または球体内の表示を繰り返し実行して、その表示位置をつないでゆくことによって、時間変化を球面上または球体内に表示することが可能となる。

## 【 0 0 4 0 】

図 1 0 は、図 5 に示した実験 1 の遺伝子発現強度データを、時間変化に従って蓄積した時系列発現データの例を示している。ここでは実験 1 に対応して配列を 1 つ用意し、一定時間間隔（1 時間ごと、1 週間ごとなど）のデータを配列の各要素に貯えるデータ構造を採っている。

## 【 0 0 4 1 】

発現データの時間変化を表示させて折れ線や曲線などを描くためには、図 1 0 の配列のインデックスを逐次変化させ、繰り返し処理によって各要素に対応するデータに貯えた蛍光色の輝度を取り出し、球面上あるいは球体内の表示位置を決めてゆき、表示位置の間を矢印や補間曲線などを描く処理をすればよい。また、複数遺伝子を前記クラスタ分析により、遺伝子グループとして捉えたい場合は、変化の状況を遺伝子グループの移動・分割・併合などとして描画処理を行えばよい。

## 【 0 0 4 2 】

表示の手順は、図 2 のフローの手順を各配列のインデックスごとに繰り返すことになる。ただし、一時点ごとのデータ処理を図 2 のフローに従って繰り返すのではなく、ステップ 2 0 0 のデータ読み込み部分やステップ 2 0 1 のデータ正規化処理など、フロー中の一部分の処理を一括して処理した後、結果表示を行うという手順をとることも可能である。

## 【 0 0 4 3 】

図 1 1 は、前記球面上の表示において、遺伝子発現データの時間変化を表示した表示例を示す図である。

表示 1 1 0 0 は、一つの遺伝子（遺伝子 1）に着目し、その遺伝子に関する発現データを逐次計算して、その時間変化を球面上の軌跡として、折れ線矢印で表示位置を繋いだものである。この例では、最初、正常細胞 A 中で強く発現していた遺伝子 1 が、疾病 C に罹患した細胞中でも、時間が経つにつれて、遅れたかたちで発現してきている状況が推測できる。

【0044】

また、表示 1 1 0 1 は、前記クラスタ分析によって計算した遺伝子クラスタ V の球面上の表示であり、図 1 1 ではこれが時間経過とともに遺伝子クラスタ V 1（1 1 0 2）と遺伝子クラスタ V 2（1 1 0 3）に分裂して 2 つのグループに分かれていった状況を、球面上の部分領域とそれを繋ぐ曲線矢印で表している。この例では、最初、特に疾病細胞 B 中で強く発現していた遺伝子グループが 2 つのグループになり、片方が正常細胞 A 中でも働き疾病細胞 C 中ではほとんど働かないもの、他方が疾病細胞 C 中でも働き正常細胞 A 中ではほとんど働かないもの、へと分かれていった状況が推測できる。

【0045】

図 1 2 は、球面上における遺伝子発現データの時間変化表示の解釈に関して、模式的に説明した図である。図 1 2 は、図 1 1 の球面表示用に用意した 8 分の 1 球面領域を、三角形の領域として展開して対応づけ、球面内の部分領域を三角形内の部分領域として捉えた球面内領域の模式図である。例えば三角形領域中の上部の菱形領域 a は、図 1 1 の球面上で Z 軸方向の最も上方にある球面内の部分領域に対応している。

【0046】

これらの三角形内の部分領域をまたがって、遺伝子または遺伝子グループの発現データ表示位置が遷移する場合、遺伝子の発現状況の時間変化として解釈し、遺伝子の働きの変化を推測することができる。例えば、最初領域 b に表示されていた遺伝子クラスタ V（1 2 0 1）が、領域 p や領域 c に遷移して遺伝子クラスタ V 1（1 2 0 2）や遺伝子クラスタ V 2（1 2 0 3）へ移動して表示された場

合、着目している遺伝子グループは、初めは特に試料Bにおける発現強度が著しかったが、試料Cにおける発現の方が次第に相対的に強くなっていった状況が推測できる。一方、遺伝子クラスタV（1201）が領域oに遷移して遺伝子クラスタV3（1204）が表示された場合、着目している遺伝子グループは、最初の試料Bにおける発現が著しかった傾向がなくなり、試料A・試料B・試料Cのいずれでも特に偏りなく遺伝子グループが働くようになっていった状況が推測できる。

## 【0047】

図13は、前記球体内の表示において、遺伝子発現データの時間変化を表示した表示例を示す図である。

表示1300は一つの遺伝子（遺伝子2）に着目し、遺伝子2に関する発現データを逐次計算して、その時間変化を球体内の軌跡として、折れ線矢印で表示位置を繋いだものである。この例では、最初正常細胞A及び疾病Bに罹患した細胞中で中程度に発現していた遺伝子2が、時間が経つにつれて、細胞Aや細胞Bでの発現は弱くなると共に、疾病Cに罹患した細胞中で、特に強く発現してきている状況が推測できる。

## 【0048】

また、1301は前記クラスタ分析によって計算した遺伝子クラスタWの球体内の表示であり、図13ではこれが時間経過とともに遺伝子クラスタW1（1302）と遺伝子クラスタW2（1303）に分裂して2つのグループに分かれていった状況を、球体内の部分領域とそれを繋ぐ曲線矢印で表している。この例では、最初特に疾病細胞C中で中程度に発現していた遺伝子グループが2つのグループになり、片方が正常細胞A中ではほとんど発現せず疾病細胞Bおよび疾病細胞C中でも発現強度は同じくらいで中程度以下のもの、他方が疾病細胞B中でより強く発現し正常細胞Aおよび疾病細胞C中ではほとんど発現しないもの、へと分かれていった状況が推測できる。

## 【0049】

これらの表示方法が有効である具体的な実験としては、例えば以下のような実験を行い、遺伝子発現レベルの時間変化を調べたい状況が考えられる。

- (1) 疾病Xの罹患状況に関し、患者の体質Aタイプ・Bタイプ・Cタイプを比較する実験
- (2) 微生物（細菌など）の活動状況に関し、株A・株B・株Cを比較する実験
- (3) ある生物種の発生過程において、頭部・胴体・四肢など、3箇所の部位における複数遺伝子の発現状況を比較する実験
- (4) 3種類の臓器（肝臓・腎臓・心臓など）における遺伝子発現の差異を検出するための実験

【0050】

本表示方式によれば、これらの実験は、例えばAとB、およびAとCの実験を時間変化に従って観測しデータ収集するだけで、BとCの比較表示が可能となり、BとCを比較するための時間経過に従った実験を省略することができるため、実験労力の削減においても効果がある。

以上述べたように、時系列に観測した試料Aと試料B、及び試料Aと試料Cの実験結果から、図11に示したような遺伝子発現データの球面上のプロット及びクラスタ分析結果の時系列の表示、もしくは図13に示したような遺伝子発現データの球体内のプロット及びクラスタ分析結果の時系列の表示ができるようになり、複数遺伝子間の時系列変化に関する遺伝子機能推測が可能となる。

【0051】

【発明の効果】

以上説明したように、本発明によれば、1種類の遺伝子発現データをもとにして、2つの実験結果から他の複数遺伝子発現データを比較し、分類状況および変化の状況を大まかに把握するために有効な球体内もしくは球面上のビジュアル表示を提供することができる。

【図面の簡単な説明】

【図1】

システム構成図。

【図2】

概略処理フローを示す図。

【図 3】

標準的遺伝子発現データ分析結果表示例（平面状）の説明図。

【図 4】

発現データの分析対象範囲の設定例の説明図。

【図 5】

遺伝子発現実験の具体的データ例を示す図。

【図 6】

遺伝子発現データの表示例を示す図。

【図 7】

遺伝子発現データの表示例（球体内）を示す図。

【図 8】

遺伝子発現データの表示例（球面上）を示す図。

【図 9】

標準的遺伝子発現データの時間変化の表示例を示す図。

【図 1 0】

遺伝子発現実験の時系列に従った具体的データ例を示す図。

【図 1 1】

遺伝子発現データの時間変化の表示例（球面上）を示す図。

【図 1 2】

遺伝子発現データの時間変化表示（球面上）の解釈に関する模式的説明図。

【図 1 3】

遺伝子発現データの時間変化の表示例（球体内）を示す図。

【符号の説明】

- 1 0 0 遺伝子発現データ
- 1 0 1 表示装置
- 1 0 2 キーボード
- 1 0 3 マウス
- 1 0 4 処理部
- 8 0 0 表示用球面

8 0 1 遺伝子クラスタ U

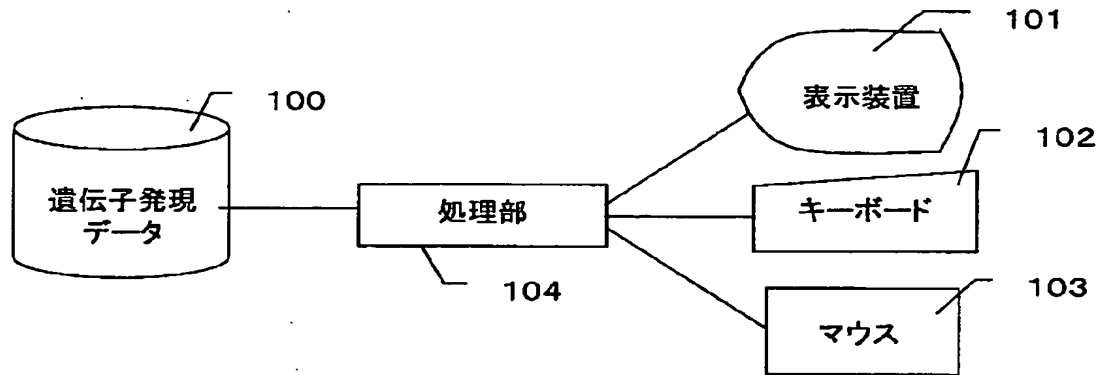
8 0 2 遺伝子クラスタ V

8 0 3 遺伝子クラスタ W

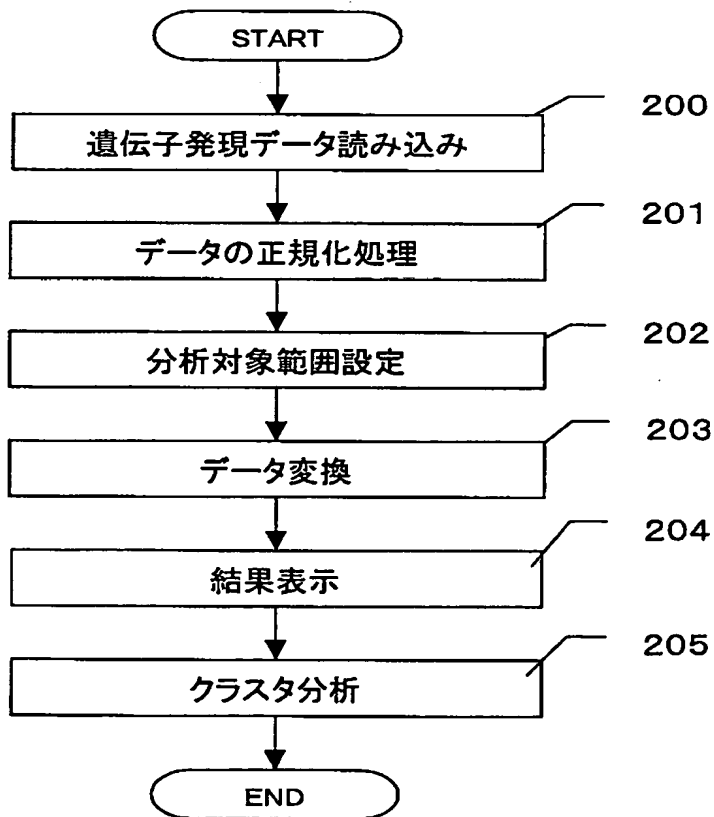


【書類名】 図面

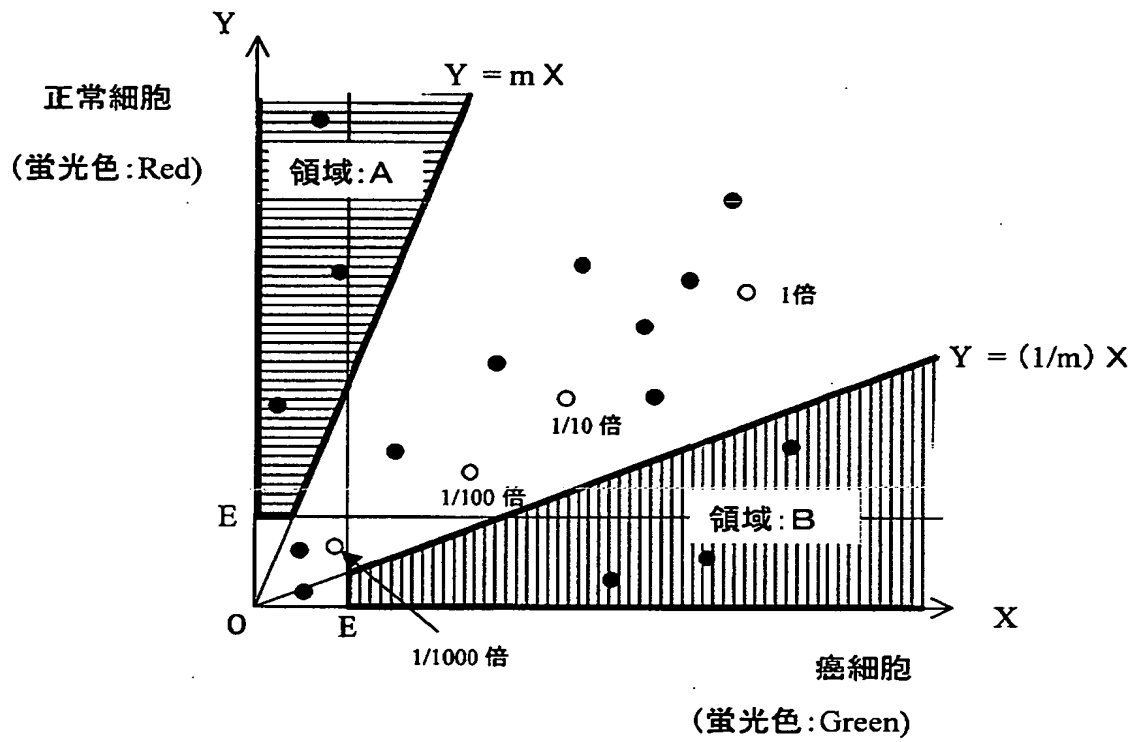
【図 1】



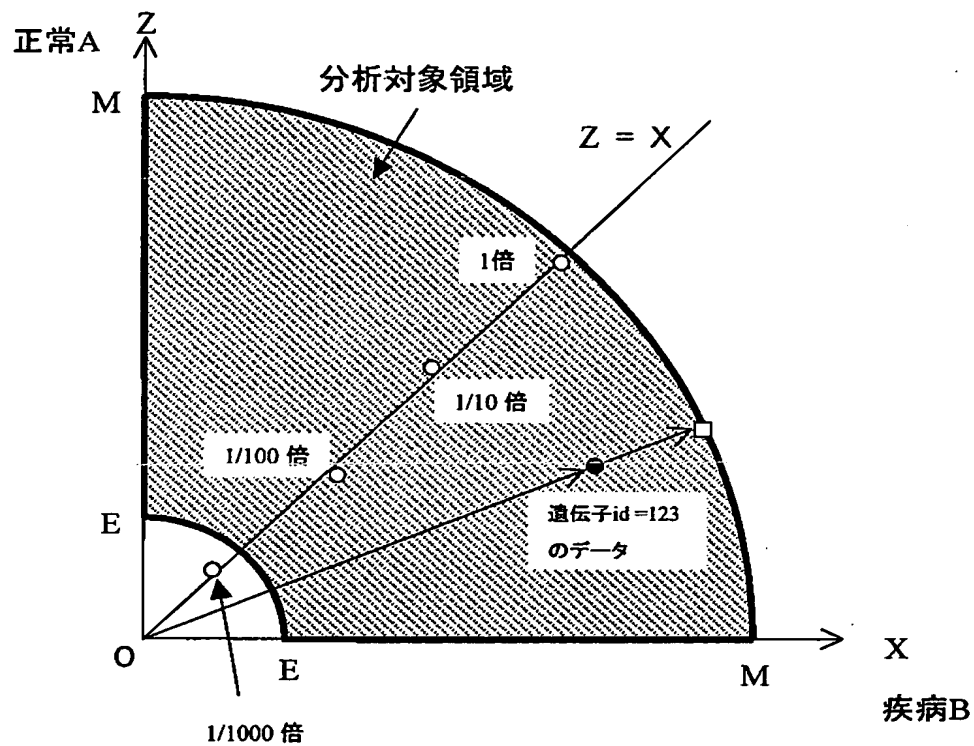
【図 2】



【図 3】



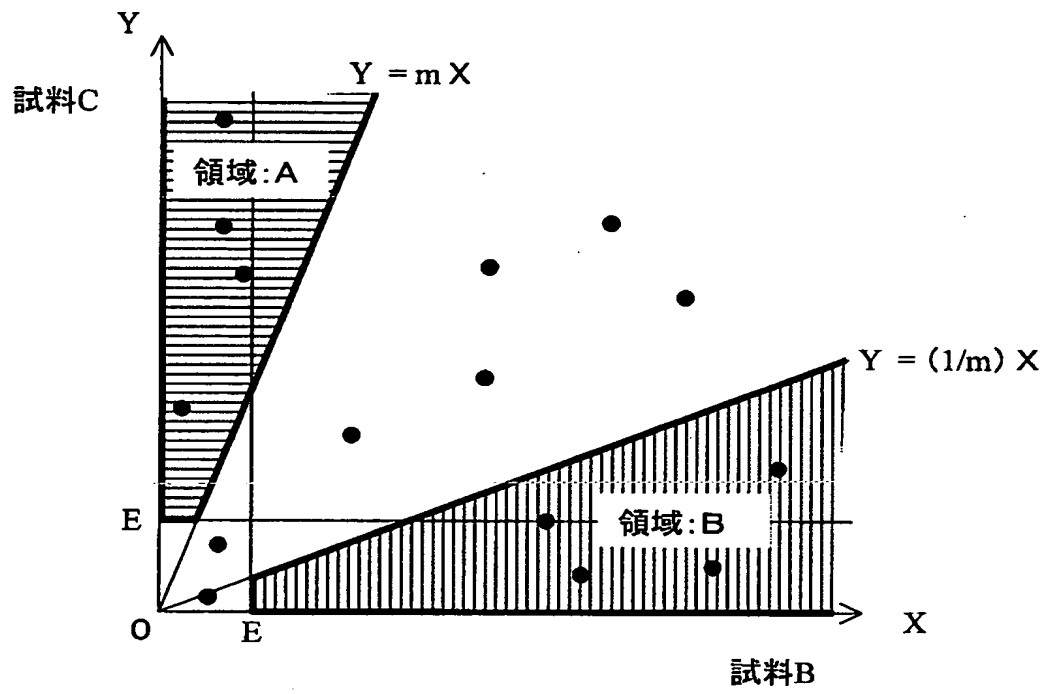
【図 4】



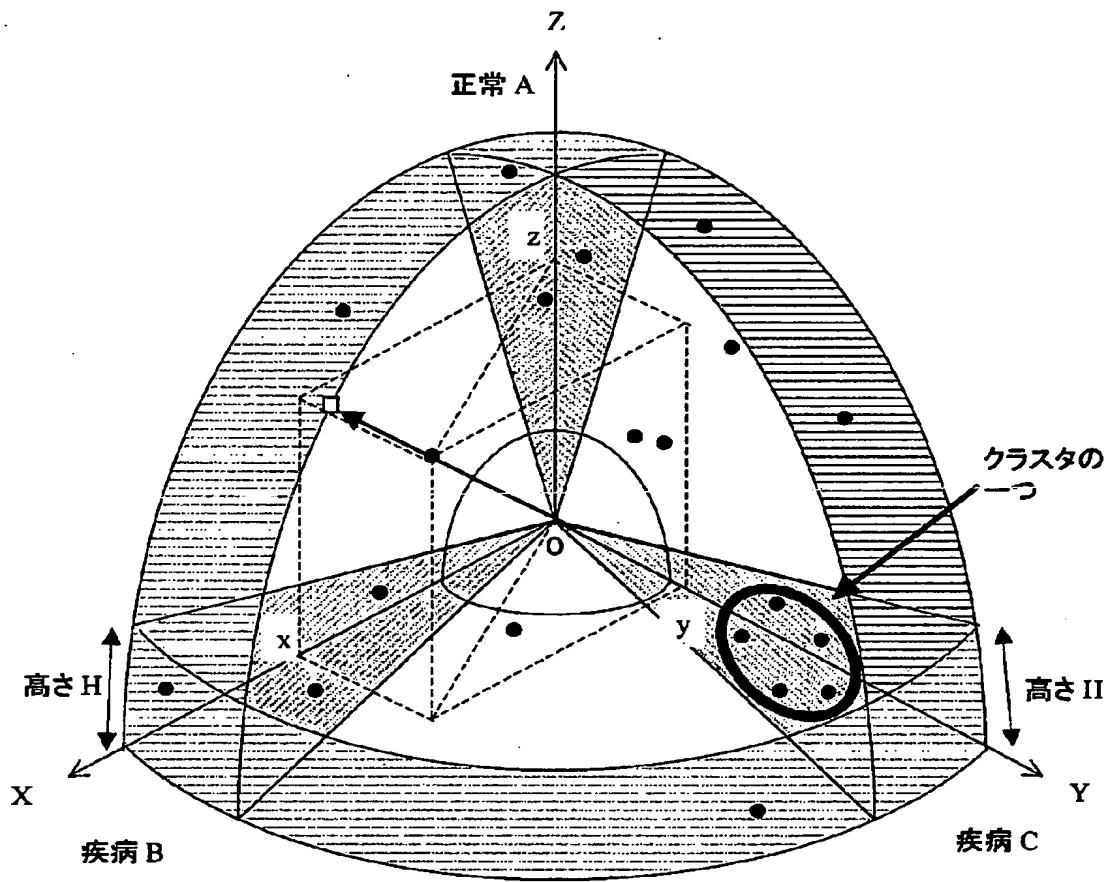
【図 5】

	実験1		実験2	
遺伝子ID	正常A	疾病B	正常A	疾病C
1	1,234	56	1,025	72
2	11,224	888	336	7,728
3	107	3,408	10,043	298
⋮	⋮	⋮	⋮	⋮
m	9,753	8,905	7,531	6,907

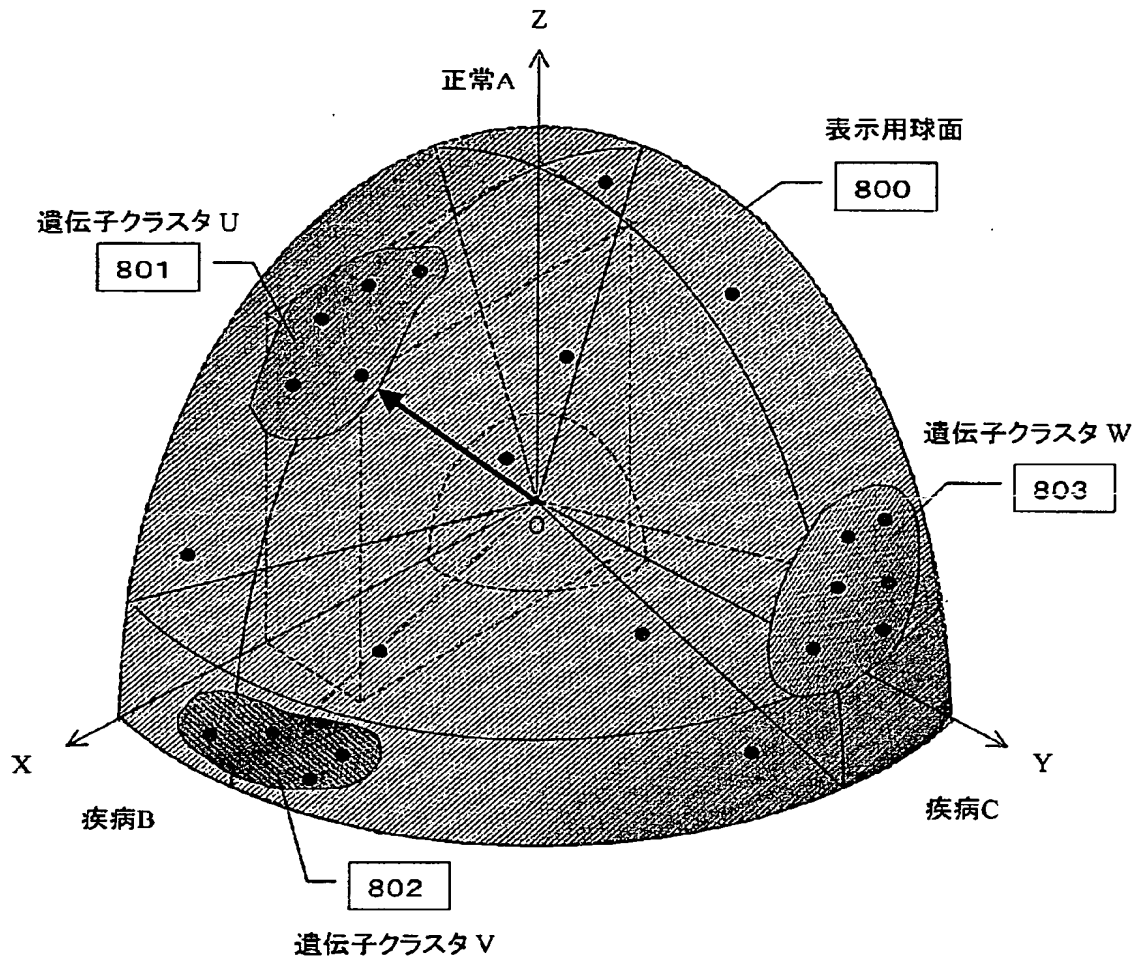
【図 6】



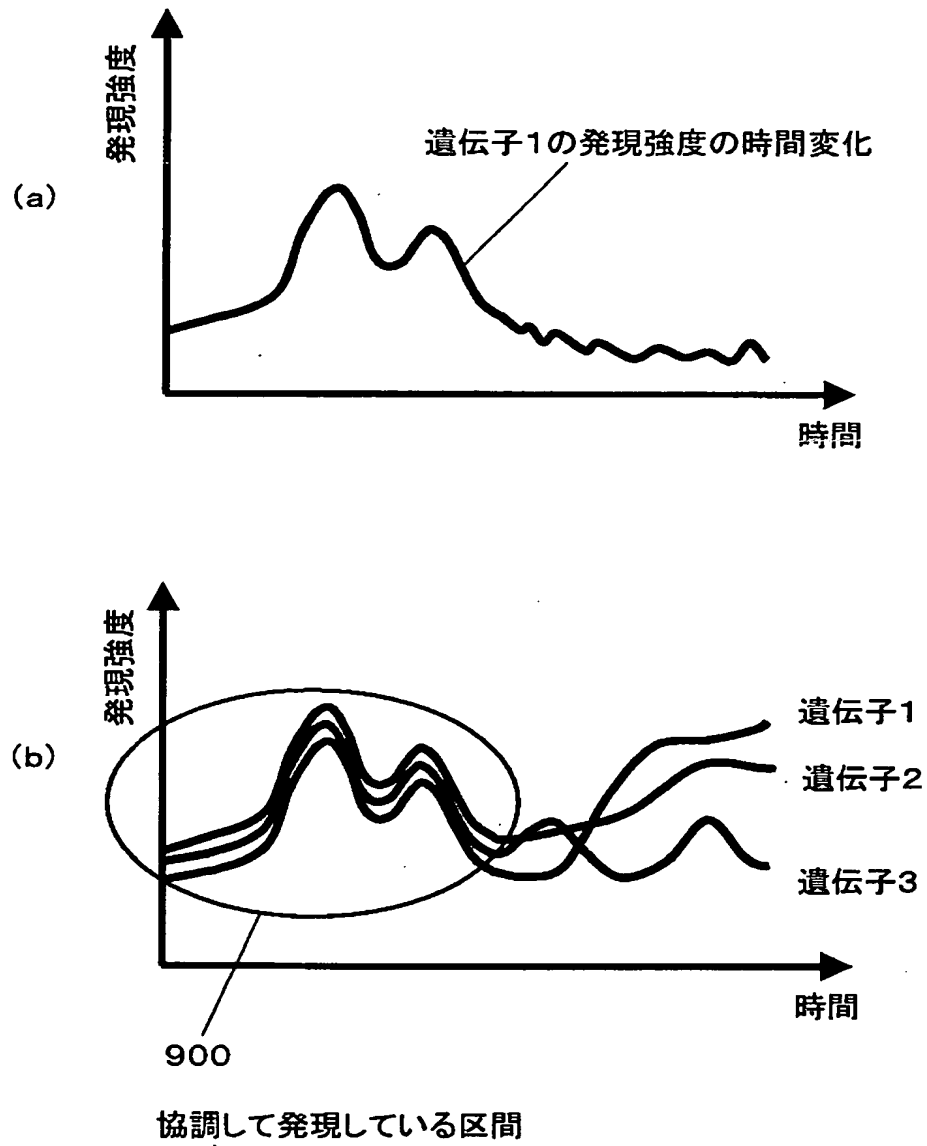
【図 7】



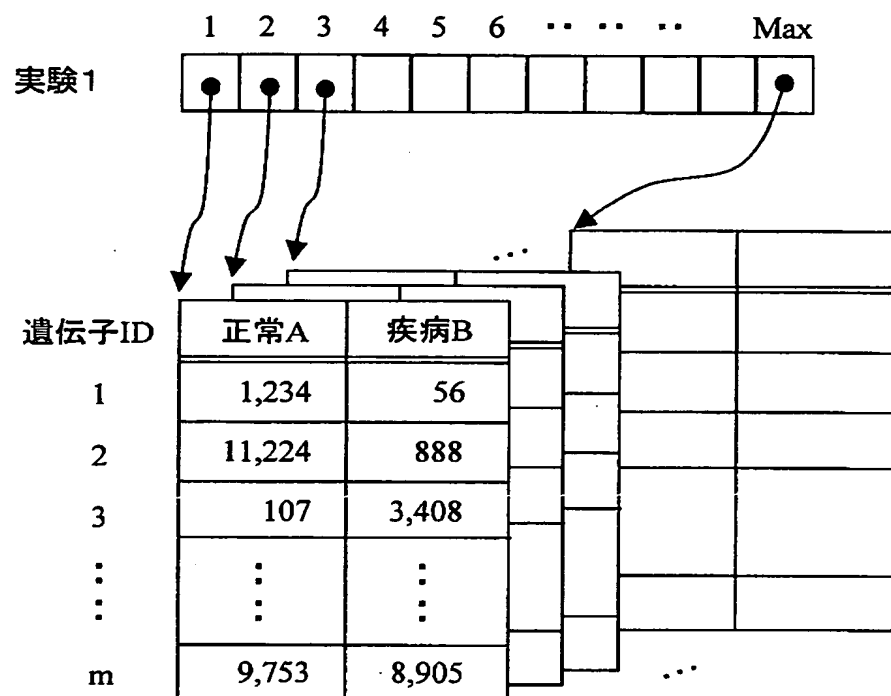
【図 8】



【図9】

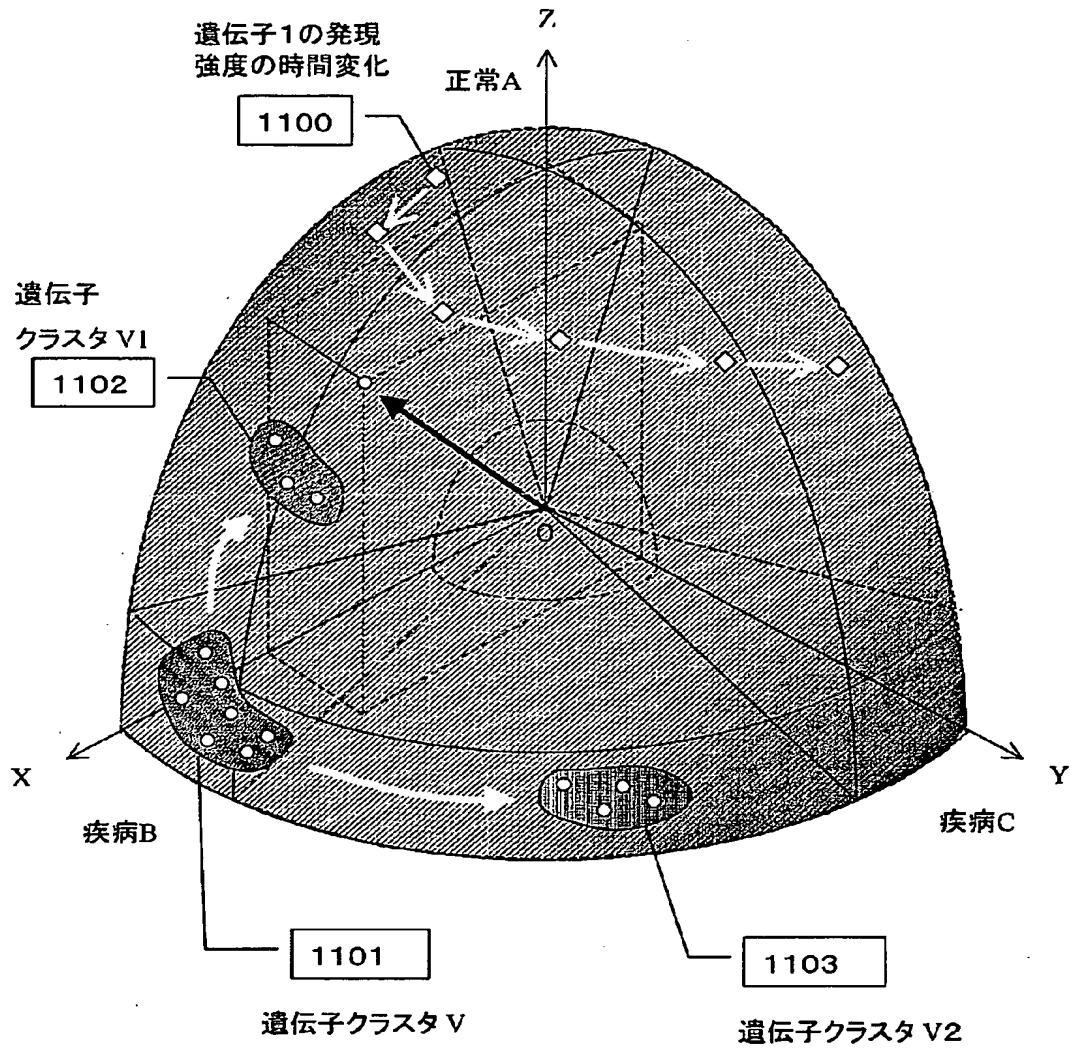


【図 1 0】

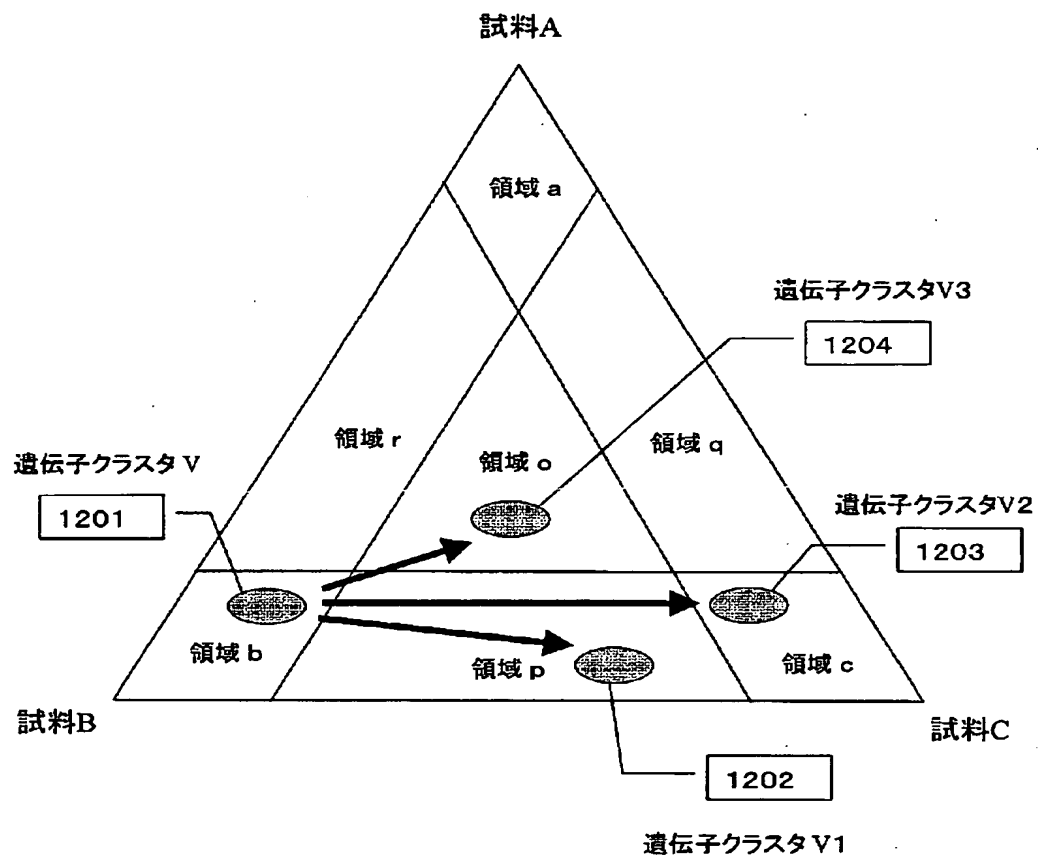




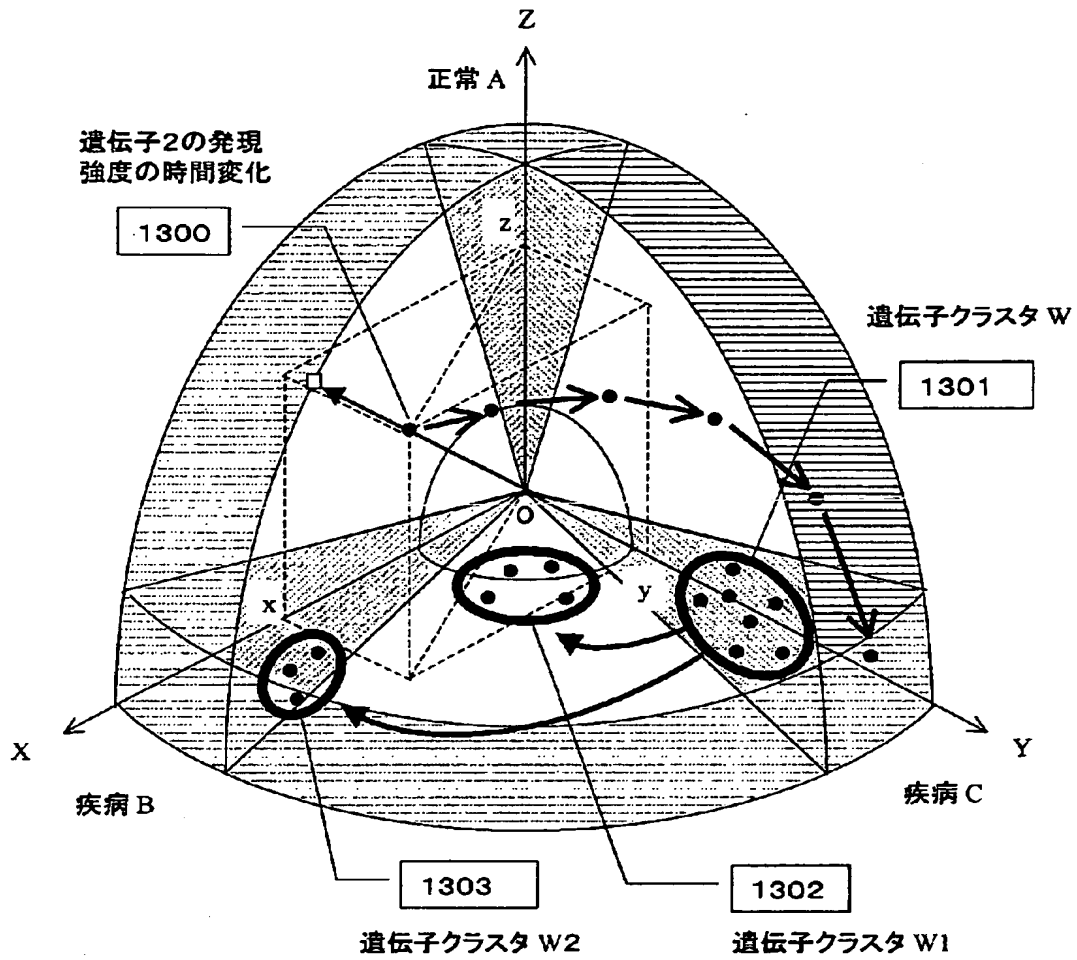
【図 11】



【図 12】



【図13】



【書類名】 要約書

【要約】

【課題】 1 種類の遺伝子発現データをもとにして、2 つの実験結果から他の複数遺伝子発現データを比較し、分類状況および変化の状況を大まかに把握するために有効なビジュアル表示を提供する。

【解決手段】 2 種類ごとの試料に対する発現強度のデータを異なる実験結果の間で比較するため、共通に使用された試料のデータを媒介として 3 種類の発現強度を 3 次元表示する。具体的には試料 A と試料 B の発現量データと試料 A と試料 C の発現量データを一つの 3 次元データに変換し、球体内の点として表示する。あるいは、試料 A と試料 B の比と試料 A と試料 C の比に着目して球面上へマッピングすることで、各遺伝子の試料 A, B, C における発現状況を球面上の分布として表示する。これら球体内の点の分布や球面上の点の分布をもとにクラスタ分析することで、3 種類の試料における遺伝子発現状況をビジュアルに把握する。また、球面上又は球体内の点として表示した発現強度データを、時間変化に従って球面上又は球体内の軌跡として、各遺伝子又はクラスタ分析による遺伝子グループに対し折れ線又は曲線を描くことにより、発現状況の時間変化をビジュアルに把握する。

【選択図】 図 8

出 願 人 履 歴 情 報

識別番号 [000233055]

1. 変更年月日 1990年 8月 7日

[変更理由] 新規登録

住 所 神奈川県横浜市中区尾上町6丁目81番地

氏 名 日立ソフトウェアエンジニアリング株式会社